Amendment to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

1. (currently amended) An audio-visual content synthesis apparatus ~~in a digital communication system that is capable of~~ for (i) receiving audio-visual input signals that represent a speaker who is speaking and ~~capable of~~ (ii) creating an animated version of the speaker's face ~~of the speaker using a plurality of audio logical units~~ that represent the speaker's speech, said apparatus comprising ~~a content synthesis application processor that~~:

means for extracting (i) ~~extracts~~ audio features of the speaker's speech and (ii) visual features of the speaker's face from the audio-visual input signals;

means for creating ~~creates~~ audiovisual input vectors from (i) the extracted audio features and (ii) the extracted visual features, wherein each audiovisual input vector comprises a hybrid logical unit that exhibits properties of both (a) the phonemes and (b) the visemes;

means for creating ~~creates~~ audiovisual configurations from the audiovisual input vectors, wherein the audiovisual configurations comprise speaking face movement components in an audiovisual space; and

means for performing ~~performs~~ a semantic association procedure on the audiovisual input vectors to obtain an association between phonemes that represent the speaker' speech and visemes that represent the speaker's face for each audiovisual input vector.

2. (currently amended) [[An]] The apparatus as claimed in Claim 1, further comprising: ~~wherein the content synthesis application processor is capable of~~

means for analyzing an input audio signal, wherein said input audio signal analyzing means is configured for: ~~by:~~

extracting audio features of a speaker's speech from the input audio signal;

finding corresponding video representations for the extracted audio features using a semantic association procedure; and

matching the corresponding video representations with the audiovisual configurations.

3. (currently amended) [[An]] The apparatus as claimed in Claim 2, further comprising: ~~wherein the content synthesis application processor is further capable of:~~

means for creating a computer generated animated face for each selected audiovisual configuration;

means for synchronizing each computer generated animated face with the speaker's speech of the input audio signal; and

means for outputting an audio-visual representation of the speaker's face synchronized with the speaker's speech.

4. (currently amended) [[An]] The apparatus as claimed in Claim 1, wherein the audio features ~~that the content synthesis application processor extracts~~ extracted from the audio-visual input signals comprise one of: Mel Cepstral Frequency Coefficients, Linear Predictive Coding Coefficients, Delta Mel Cepstral Frequency Coefficients, Delta Linear Predictive Coding Coefficients, and Autocorrelation Mel Cepstral Frequency Coefficients.

5. (currently amended) [[An]] The apparatus as claimed in Claim 1, wherein said ~~content synthesis application processor~~ means for creating audiovisual configurations creates the audiovisual configurations from the audiovisual input vectors using one of: a Hidden Markov Model and a Time Delayed Neural Network.

6. (currently amended) [[An]] The apparatus as claimed in Claim 2, wherein said ~~content~~

~~synthesis application processor~~ analyzing matches the corresponding video representations with the audiovisual configurations using one of: a Hidden Markov Model and a Time Delayed Neural Network.

7. (currently amended) [[An]] The apparatus as claimed in Claim 3, further comprising: ~~wherein said content synthesis application processor further comprises:~~

means for implementing a facial audio visual feature matching and classification module that matches each of a plurality of audiovisual configurations with a corresponding classified audio feature to create a facial animation parameter; and

means for implementing a facial animation for selected parameters module that creates an animated version of the face of the speaker for a selected facial animation parameter.

8. (currently amended) [[An]] The apparatus as claimed in Claim 7, wherein said facial animation for selected parameters module creates an animated version of the face of the speaker by using one of: (1) 3D models with texture mapping and (2) video editing.

9. (currently amended) [[An]] The apparatus as claimed in Claim 2, wherein said semantic association procedure comprises one of: latent semantic indexing, canonical correlation, and cross modal factor analysis.

10. (canceled)

11. (currently amended) [[An]] The apparatus as claimed in Claim 8, further comprising: ~~wherein said content synthesis application processor further comprises:~~

means for implementing a speaking face animation and synchronization module that synchronizes each animated version of the face of the speaker with the audio features of the speaker's speech to create an audio-visual representation of the

speaker's face that is synchronized with the speaker's speech; and

means for implementing an audio expression classification module that determines a level of audio expression of the speaker's speech and provides said level of audio expression of the speaker's speech to said speaking face animation and synchronization module to use to modify animated facial parameters of the speaker in response to the determined level of audio expression.

12. (currently amended) A method for use in synthesizing audio-visual content in a video image processor, said method comprising the steps of:

receiving audio-visual input signals that represent a speaker who is speaking;

extracting (i) audio features of the speaker's speech and (ii) visual features of the speaker's face from the audio-input signals;

creating audiovisual input vectors from (i) the extracted audio features and (ii) the extracted visual features, wherein each audiovisual input vector comprises a hybrid logical unit that exhibits properties of both (a) the phonemes and (b) the visemes;

creating audiovisual configurations from the audiovisual input vectors, wherein the audiovisual configurations comprise speaking face movement components in an audiovisual space; and

performing a semantic association procedure on the audiovisual input vectors to obtain an association between phonemes that represent the speaker' speech and visemes that represent the speaker's face for each audiovisual input vector.

13. (currently amended) The method as claimed in Claim 12, further comprising: ~~the steps of:~~

analyzing an input audio signal of a speaker's speech, wherein analyzing includes:

extracting audio features of the speaker's speech from the input audio signal;

finding corresponding video representations for the extracted audio features

using a semantic association procedure; and

matching the corresponding video representations with the audiovisual configurations.

14. (currently amended) The method as claimed in Claim 13, further comprising the steps of:

creating a computer generated animated face for each selected audiovisual configuration;

synchronizing each computer generated animated face with the speaker's speech of the input audio signal; and

outputting an audio-visual representation of the speaker's face synchronized with the speaker's speech.

15. (currently amended) The method as claimed in Claim 12, wherein the audio features ~~that are~~ extracted from the audio-visual input signals comprise one of: Mel Cepstral Frequency Coefficients, Linear Predictive Coding Coefficients, Delta Mel Cepstral Frequency Coefficients, Delta Linear Predictive Coding Coefficients, and Autocorrelation Mel Cepstral Frequency Coefficients.

16. (currently amended) The method as claimed in Claim 12, wherein the audiovisual configurations are created from the audiovisual input vectors using one of: a Hidden Markov Model and a Time Delayed Neural Network.

17. (currently amended) The method as claimed in Claim 13, wherein the corresponding video representations are matched with the audiovisual configurations using one of: a Hidden Markov Model and a Time Delayed Neural Network.

18. (currently amended) The method as claimed in Claim 12, further comprising the

steps of:

matching each of a plurality of audiovisual configurations with a corresponding classified audio feature to create a facial animation parameter; and

creating an animated version of the face of the speaker for a selected facial animation parameter.

19. (currently amended) The method as claimed in 18, further comprising the step of:

creating an animated version of the face of the speaker by using one of: (1) 3D models with texture mapping and (2) video editing.

20. (currently amended) The method as claimed in Claim 13, wherein said semantic association procedure comprises one of: latent semantic indexing, canonical correlation, and cross modal factor analysis.

21. (canceled)

22. (currently amended) The method as claimed in Claim 20, further comprising the steps of:

synchronizing each animated version of the face of the speaker with the audio features of the speaker's speech;

creating an audio-visual representation of the face of the speaker that is synchronized with the speaker's speech;

determining a level of audio expression of the speaker's speech; and

modifying animated facial parameters of the speaker in response to a determination of the level of audio expression of the speaker's speech in response to the determined level of audio expression.

23. – 33. (canceled)